

# Algorithm for overlapping estimation of common change-sets in spatial data of fixed size

Leonid Torgovitski<sup>\*†</sup>

*Mathematical Institute, University of Cologne  
Weyertal 86-90, 50931, Cologne, Germany*

## Abstract

We propose a flexible class of estimates for “common change in the mean” sets in spatio-temporal data. We rely on a scan type approach by subdividing the spatial observations into suitable overlapping regions to which classical CUSUM (cumulative sums) estimates may then be applied separately. The aggregated “local” estimates are used to construct consistent “global” estimates of the change set(s) by taking the overlapping structure into account. The domain and the change regions may have irregular shapes and the suggested procedure is especially suited for estimation of multiple change regions. The performance is demonstrated in a simulation study.

## Keywords

Change-set estimation, Spatio-temporal data, Fixed time, Common change-sets, Multiple changes

## 1 Introduction

The estimation of common change points (e.g. points which correspond to changes of the mean) in the framework of multivariate time series  $X_1, X_2, \dots, X_N \in \mathbb{R}^d$ , with fixed sample size  $N$  and an increasing dimension  $d \rightarrow \infty$ , received some attention in recent literature (cf., e.g., Bai (2010), Bleakley and Vert (2010), Hadri et al (2012), Kim (2014) and Torgovitski (2015)). Within such a setting, Bai (2010) studied a *single common change point* model and considered a classical least squares estimate, whereas Bleakley and Vert (2010, 2011a) considered a *multiple* common change point model and adapted the *total variation denoising* approach to it. As pointed out in Torgovitski (2015), both methods may be seen as special cases of some general class of weighted CUSUM (cumulative sums) estimates. In the latter article, the consistency is studied for a whole class of such estimates which will also play a major role in the present work.

In this paper we turn to time series of spatial data  $X_1, X_2, \dots, X_d \in \mathbb{R}^{m \times n}$ , where the parameters  $m, n$  are fixed and  $d \rightarrow \infty$ . (For our approach, the time parameter  $d$  will turn out to correspond to the previously mentioned dimension parameter.) Our aim is to develop an algorithmic framework for the estimation of *change sets*  $S$ , where the means  $E(X_k(i, j))$  for  $(i, j) \in S$  differ from the means

---

<sup>\*</sup>E-mail: ltorgovi@math.uni-koeln.de

<sup>†</sup>Research partially supported by the Friedrich Ebert Foundation.

corresponding to  $(i, j) \notin S$ . The proposed algorithm is based on theoretical results in the aforementioned works of Bai (2010), Bleakley and Vert (2011a) and Torgovitski (2015). The method is especially suitable for the estimation of multiple change sets, where their number does not necessarily has to be known in advance, and the approach may also be easily extended to other, more complex situations, e.g., with an irregular domain and irregular change sets. As a special case, the same principle may also be applied in a straightforward manner to the common multiple change point estimation within the panel data framework in Torgovitski (2015) or Bleakley and Vert (2011a) in order to obtain consistent estimates for all changes when the number of panels tends to infinity.

For demonstration purposes, the suggested approaches are implemented with a graphical user interface as a Matlab application which can be obtained from the author or via [www.mi.uni-koeln.de/~ltorgovi](http://www.mi.uni-koeln.de/~ltorgovi).

For some problems and approaches for change-set detection that are remotely related to our situation under consideration we refer the reader e.g. to Arnold and Wied (2012), Arnold et al (2014) and the references therein (in particular to Polzehl and Spokoiny (2000) and to the review article of Qui (2007)).

## Notation

First, we need to introduce some notation in order to formulate our model. Consider a set  $S \subset \mathbb{R}^2$  of two-dimensional points, i.e.

$$S = \{(i, j) \mid i = i_1, \dots, i_m \in \mathbb{N}, j = j_1, \dots, j_n \in \mathbb{N}, m, n \in \mathbb{N}\}.$$

We call any point  $(i, j \pm 1), (i \pm 1, j) \in S$  to be *adjacent* or a *neighbour* to  $(i, j) \in S$ . Correspondingly, two sets  $S_1, S_2 \subset S$  are called adjacent if at least two nodes  $u \in S_1, v \in S_2$  exist that are adjacent to each other. Furthermore, we will associate  $S$  with an undirected graph such that each point  $(i, j) \in S$  corresponds to a node and such that all adjacent nodes are connected by edges. The *boundary* of  $S$  is a subset  $B \subset S$  which contains only nodes  $(i, j) \in S$  that have less than four distinct neighbours, i.e. nodes that are not *4-connected*. Correspondingly, an *interior* node  $(i, j) \in S$  has to have four neighbours within  $S$ . The set  $S$  is called *connected* whenever the associated graph is connected, i.e., if there exists a path between any two nodes  $n_1, n_2 \in S$ .

As usual, we define the distance  $d(u, v)$  of two nodes  $u, v \in S$  w.r.t. the set  $S$  as the shortest path between them (within  $S$ ). Accordingly, we define the distance of two sets  $S_1, S_2 \subset S$  w.r.t. the set  $S$  as

$$d(S_1, S_2) = \inf\{d(u, v) \mid u \in S_1, v \in S_2\},$$

with  $\inf \emptyset = \infty$  in which case the sets are obviously disjoint. The *Jaccard distance* between two sets  $A, B \subset S$  is defined by

$$d_J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}. \quad (1.1)$$

This is a common measure of distance between two subsets, which will be used to quantify the precision of our estimates later on. We are now in a position to formulate our actual model.

## Statistical model

We consider a spatio temporal *signal plus noise* model given by

$$X_k(i, j) = m_k(i, j) + \varepsilon_k(i, j) \quad (1.2)$$

for  $k = 1, \dots, d$  and  $(i, j) \in D$ , where  $D$  is assumed to be a rectangular domain, i.e.

$$D = \{(i, j) \mid i = 1, \dots, m, j = 1, \dots, n\},$$

with  $m, n \geq 4$  being fixed integers. Here, the  $m_k(i, j)$  are the deterministic signals and  $\varepsilon_k(i, j)$  are the random variables representing the noise. One may interpret the sequence  $\{X_k\}$  as a random field defined on the lattice  $D$  and consider  $k$  to be the time parameter. Throughout, we assume the family of random variables

$$\{\varepsilon_k(i, j), i = 1, \dots, m, j = 1, \dots, n\}$$

to be i.i.d. for each  $k$ , identically distributed in  $k$  with  $0 < E(\varepsilon_1(1, 1)^2) = \sigma^2 < \infty$  and centered. The data may be dependent in time, for which appropriate conditions will be imposed later on. Additionally, we have to assume uniformly bounded finite fourth moments.

For the general setting we may assume a partitioning of the domain  $D$  as

$$D = S_1 + \dots + S_K, \quad (1.3)$$

for some  $K > 1$ , where each set  $S_k$ ,  $k = 1, \dots, K$  is non-empty and is assumed to be connected. Further, we assume piecewise constant means, i.e.

$$m_k(i, j) = \sum_{l=1}^K m_k(S_l) \mathbb{1}_{S_l}(i, j), \quad (1.4)$$

with  $m_k(S_l) \in \mathbb{R}$ , such that for any  $k$  it holds that  $m_k(S_l) \neq m_k(S_u)$  for all adjacent sets  $S_l, S_u$  with  $u \neq l$ . Our goal is the estimation of the partition  $S_1, \dots, S_K$  based on the noisy sequence  $X_k$ . (The Assumption (1.4) can be related e.g. to Polzehl and Spokoiny (2000, eq. (2)) but the statistical model there is non temporal.) As already mentioned, for  $m = 1$  the setting (1.2) and (1.4) fits into the multiple change point scenario in panel data with fixed time parameter where the number of panels  $d$  tends to infinity (cf., eg., Bleakley and Vert (2011a) and Torgovitski (2015)).

The rectangular domain  $D$  is chosen for simplicity of exposition only and as already mentioned all results discussed in this paper can be easily extended to more complex situations in a straightforward manner. Recall that we consider asymptotics for  $d \rightarrow \infty$ .

**Definition 1.1.** Assume the partitioning (1.3). We will call the sets  $S_l$  for  $l = 1, \dots, K$  to be *common change sets* if:

1. The sets  $S_l$  are connected.
2. For all *total average changes*, defined as

$$\Delta_\infty^2(S_l, S_p) = \lim_{d \rightarrow \infty} \frac{\sum_{k=1}^d |\Delta_k|^2}{d}, \quad (1.5)$$

with  $\Delta_k(S_l, S_p) = m_k(S_l) - m_k(S_p)$ , it holds, for all adjacent sets  $S_l, S_p$  with  $l \neq p$ , that

$$0 < \Delta_\infty^2(S_l, S_p) < \infty, \quad (1.6)$$

which quantifies the notion “common” in our setting.

Since the extension to multiple change sets is straightforward, we will (mostly) restrict ourselves in Sections 2-4 below to the single change set case, i.e. to  $K = 2$ , where  $D = S + S^c$  and  $S, S^c$  are both formally common change sets. However, here it is more convenient to think that  $S^c$  reflects the normal state region and  $S$  is the only common change set differing from that normal state. We will use this terminology for brevity and simply write  $\Delta_\infty^2 = \Delta_\infty^2(S, S^c)$  in this situation.

## Motivation

The model (1.2)-(1.4) with change sets  $S_1, S_2, \dots$  states a natural spatial extension of the setting considered e.g. by Torgovitski (2015).

One may think of digital imaging and assume a rectangular image sensor, i.e. an array of pixel sensors, corresponding to the domain  $D$ . Further, assume the image sensor to record a large test-sequence  $X_1, \dots, X_d$  of images that represent the light intensity (e.g. as monochrome grayscale images) and assume that the measurements, i.e. the images, are affected by some random noise.

Altogether, each image corresponds to an observation  $X_k$  where  $X_k(i, j)$  represents the measured intensity by the  $(i, j)$ -th pixel, where  $m_k(i, j)$  is the true image intensity (cf., e.g., Qui (2007)). Now, one may think of change sets  $S_1, S_2, \dots$  to correspond e.g. to objects in the image that should be segmented or to a set of faulty pixels. The estimates discussed in this article can be used to estimate such sets based on a sufficiently long sequence of observations, i.e. when  $d$  is large.

In contrast to more established settings and approaches (cf., e.g., Polzehl and Spokoiny (2000) and Qui (2007)) our aim here is to identify the partitioning based on a whole sequence of images. It is important that our model allows the means  $m_k$  to change in each observation  $k$  at any point  $(i, j)$ . For instance, we may think of changing lighting conditions while the  $X_k, k = 1, \dots, d$  are recorded. (Otherwise, as shown in Figure 5, assuming the means  $m_k(i, j)$  to be constant across all  $k$ , one could e.g. simply rely on averages  $\bar{X}_k(i, j)$  for each point  $(i, j)$  to obtain a partitioning.)

This article is organized as follows. We begin with preliminaries in Section 2, where we briefly recall some theoretical results of Bleakley and Vert (2011a) and Torgovitski (2015) on which our algorithms will be based. In Section 3 we describe the estimation algorithm together with the conditions that ensure consistency of the estimates. In Section 4 we show some simulation results to demonstrate the performance for finite  $d$  and especially to show that even moderate  $d$ 's yield reasonable results.

## 2 Preliminaries

Assume a “single” common change set scenario, i.e., observations  $X_k$  in (1.2) on a domain  $D = S + S^c$  with common change sets  $S, S^c$ . Further, consider  $(i, r)$ -th horizontal *sub-slices*

$$\mathbf{Y}_j^{(i,r)} = [Y_{j,1}^{(i,r)}, \dots, Y_{j,d}^{(i,r)}]^T, \quad j = 1, \dots, N, \quad (2.1)$$

with  $Y_{1,k}^{(i,r)}, \dots, Y_{N,k}^{(i,r)}$ , obtained from these observations  $X_k$  by setting

$$Y_{j,k}^{(i,r)} := X_k(i, r + j - 1)$$

for each  $k = 1, \dots, d, j = 1, \dots, N$  and some  $r \in \{1, \dots, n - N + 1\}$  with  $4 \leq N \leq n$ . Accordingly, we set  $\eta_{j,k}^{(i,r)} := \varepsilon_k(i, r + j - 1)$  for the corresponding innovations. The series  $\mathbf{Y}_1, \dots, \mathbf{Y}_N$  can be interpreted as panel data with  $d$  panels and finite time horizon  $N$ . Notice, that the time parameter of the original spatial observations has now become a dimension parameter.

Assume that the particular  $(i, r)$ -th sub-slice  $\{\mathbf{Y}_j^{(i,r)}\}_{j=1, \dots, N}$  intersects a change set  $S$  such that

$$(i, r + j - 1) \in \begin{cases} S^c & \text{for } j = 1, \dots, u(i, r), \\ S & \text{for } j = u(i, r) + 1, \dots, N, \end{cases} \quad (2.2)$$

holds true for some  $1 \leq u(i, r) < N$  and where  $r \in \{1, \dots, n - N + 1\}$ . This situation is illustrated for the  $(9, 3)$ -rd sub-slice of size  $N = 6$  and with  $u(9, 3) = 2$  in Figure 1 below. (We set formally  $u(i, r) = \infty$  whenever (2.2) does not hold.)

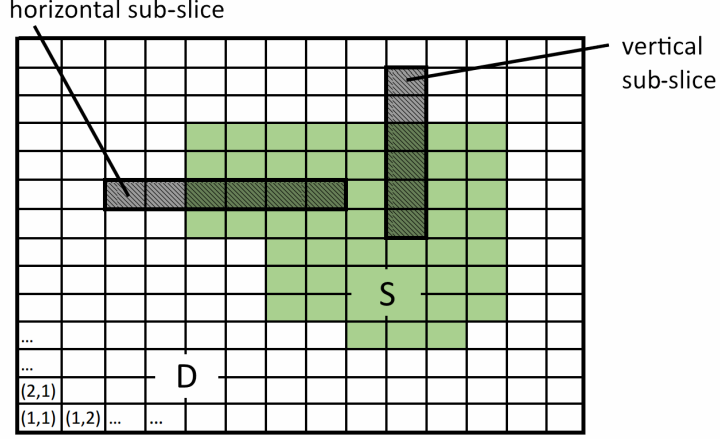


Figure 1: The scheme represents the domain  $D$  and sub-slices which intersect the change set  $S$ .

Altogether, we have a classical *change in the mean* scenario for any  $\{\mathbf{Y}_j^{(i,r)}\}_{j=1,\dots,N}$  where  $u = u(i, r)$  corresponds to a single change point, i.e. (2.1), fits into the framework considered in Torgovitski (2015). In order to estimate  $u$  one may use a weighted CUSUM estimate

$$\hat{u} = \hat{u}(i, r) = \arg \max_{p=1,\dots,N-1} w(p, N) \sqrt{\sum_{k=1}^d \left| \sum_{j=1}^p (Y_{j,k}^{(i,r)} - \bar{Y}_{N,k}^{(i,r)}) \right|^2}. \quad (2.3)$$

Here, we restrict our considerations to a typical class of weighting functions, i.e.

$$w(p, N) = ((p/N)(1 - p/N))^{-\gamma}$$

parametrized by some  $\gamma \in [0, 1/2)$  which controls the sensitivity. The  $\arg \max$  in (2.3) is defined, as usual, as the smallest index at which the maximum is attained and  $\bar{Y}_{N,k} = \sum_{i=1}^N Y_{i,k}/N$ .

To define classes of reasonable estimates for our original model (1.2) we would like to make use of the sub-slices (2.1) and of the estimates (2.3), together with the corresponding theoretical results of Bleakley and Vert (2011a) and Torgovitski (2015). Therefore, we define the normalized *noise to change ratio* parameter  $\rho$  w.r.t. the set  $S$  and w.r.t. to the length of sub-slices  $N$  as

$$\rho(S) = \frac{1}{N} \frac{\sigma^2}{\Delta_\infty^2}. \quad (2.4)$$

We also need the following assumption corresponding to Torgovitski (2015, Assumption (2.14)):

$$\frac{1}{d^2} \text{Var} \left( \sum_{k=1}^d \left( \sum_{j=1}^p (\eta_{j,k} - \bar{\eta}_{N,k}) \right)^2 \right) = o(1), \quad (2.5)$$

as  $d \rightarrow \infty$ , for every  $p = 1, \dots, N-1$ . (2.5) is a weak dependence condition that e.g. clearly holds true if the  $X_k$  are  $m$ -dependent (in particular independent) and identically distributed in  $k$  (cf. Torgovitski (2015)). Notice that we write  $\eta_{j,k}$  instead of  $\eta_{j,k}^{(i,r)}$  because (2.5) does not depend on the parameters  $i, r$ .

Our starting point is the fact that the above estimates (2.3) consistently estimate  $u$  under the assumptions on the model (1.2) and under the Assumptions (2.2) and (2.5). More precisely, we rely on the following “switching” behaviour of the estimates:

1. If there is a single change point  $u(i, r)$  w.r.t. the  $(i, r)$ -th sub-slice (2.1), such that condition (2.2) is fulfilled, then  $\hat{u}$  estimates the true change-point consistently for any  $\gamma \in [0, 1/2)$  if the noise to change ratio is below a positive threshold  $R(\gamma, u, N)$ . As  $d \rightarrow \infty$ , it holds that

$$P(\hat{u} = u) \rightarrow 1 \quad (2.6)$$

for any  $\gamma \in [0, 1/2)$  given that  $\rho < R(\gamma, u, N)$  (cf., eg., Bleakley and Vert (2011a, Theorem 2) and Torgovitski (2015, Theorems 2.6 and 2.13)). The optimal threshold  $R$  strongly depends on the parameter  $\gamma$ . The particular values for  $R(0, u, N)$  may be obtained from Bleakley and Vert (2011a, Theorem 2) in a closed form. Furthermore, it holds that  $R(1/2, u, N) = \infty$  (cf., e.g., Bleakley and Vert (2011a, Theorem 3) in the Gaussian i.i.d. case and take Torgovitski (2015, Theorem 2.6) into account regarding the nonparametric and dependent settings). We set

$$\mathcal{R}(\gamma, N) := \min_{u=1, \dots, N-1} R(\gamma, u, N),$$

which again does not depend on parameters  $i, r$ .

2. If there is no change point at all in the  $(i, r)$ -th sub-slice (2.1), i.e. if it holds that  $(i, r + j - 1) \notin S$  for  $j = 1, \dots, N$  or that  $(i, r + j - 1) \in S$  for  $j = 1, \dots, N$ , then  $\hat{u}$  estimates a spurious change. It holds that, as  $d \rightarrow \infty$ ,

$$P(\hat{u} = \lfloor N/2 \rfloor \vee \hat{u} = \lceil N/2 \rceil) \rightarrow 1 \quad (2.7)$$

for any  $\gamma \in [0, 1/2)$  (cf. Torgovitski (2015, Remark 2.7)).

We do not have closed form expressions for  $R(\gamma)$  if  $\gamma \in (0, 1/2)$ . However, from Torgovitski (2015, disp. (2.12)) it is clear that  $R(\gamma)$  tends to infinity as  $\gamma \uparrow 1/2$  (cf. also further approximations to  $R(\gamma)$  in Torgovitski (2015, Proposition 2.15)).

The above switching behaviour in (2.6) and in (2.7) will provide consistent change set estimates  $\hat{S}$  for  $S$  in Section 3. Notice that the above switching property holds true for series (2.1) of any length  $N \geq 4$ , i.e. also for small single digit series of size  $N \in \{4, \dots, 9\}$ . In order to have a unique limit in (2.7) we will consider only even  $N$ . Finally, we would like to mention that any other estimate  $\hat{u}$  with an analogous switching behaviour might be used for scanning and aggregation in the next section as well.

### 3 Estimation procedure

We stick to the single common change set scenario of the previous section. The key to the estimation of change sets in model (1.2) will be a horizontal and/or a vertical overlapping scanning approach. The idea is to reduce the global problem of the change set estimation to many local single change point problems. This will allow us to lean on the results of Bleakley and Vert (2011a) and of Torgovitski (2015) which were summarized in Section 2.

We propose a four step procedure which is outlined in the following. Notice that each step 1-4 may be performed horizontally or vertically even though some steps are described for the horizontal approach only. Moreover, we explicitly allow to combine the horizontal with the vertical approach by proceeding consecutively. (The vertical approach proceeds in the very same manner with the obvious modifications. Clearly, the notation of the previous Section 2 has to be adapted as well which will also be indicated below.)

#### 1. Slicing:

- The time series  $\{X_k\}$  is sliced into  $m$  non-spatial  $d$ -dimensional time series  $\{Y_{j,k}^{(i)}, k = 1, \dots, d\}_{j=1, \dots, n}$  for  $i = 1, \dots, m$  given by

$$Y_{j,k}^{(i)} := X_k(i, j), \quad j = 1, \dots, n, \quad k = 1, \dots, d. \quad (3.1)$$

We will denote  $Y_{j,k}^{(i)}$  as a the  $i$ -th horizontal *slice* in the following. Similarly, we may define a  $j$ -th vertical slice by  $\tilde{Y}_{i,k}^{(j)} := X_k(i, j)$  for any  $i = 1, \dots, m$ ,  $k = 1, \dots, d$ .

- Now, we tacitly assume that  $N$  is even and subdivide each  $i$ -th horizontal slice into  $n - N$  overlapping sub-slices

$$\mathbf{Y}_j^{(i,r)} = [Y_{j,1}^{(i,r)}, \dots, Y_{j,d}^{(i,r)}]^T, \quad j = 1, \dots, N \quad (3.2)$$

of size  $N$ , which are indicated by the parameter  $r$ , and where

$$Y_{j,k}^{(i,r)} := Y_{r+j-1,k}^{(i)}, \quad j = 1, \dots, N \quad (3.3)$$

for  $k = 1, \dots, d$ ,  $r = 1, \dots, n - N + 1$  and  $i = 1, \dots, m$ . Similarly, we may subdivide the  $j$ -th vertical slice into vertical sub-slices by

$$\tilde{\mathbf{Y}}_i^{(j,r)} = [\tilde{Y}_{i,1}^{(j,r)}, \dots, \tilde{Y}_{i,d}^{(j,r)}]^T, \quad i = 1, \dots, N,$$

where  $\tilde{Y}_{i,k}^{(j,r)} := \tilde{Y}_{r+i-1,k}^{(j)}$ ,  $i = 1, \dots, N$ , for  $k = 1, \dots, d$ ,  $r = 1, \dots, n - N + 1$  and  $j = 1, \dots, n$ . Definition (3.2) resembles (2.1). However, in the 3rd step it will be convenient to think of all horizontal (vertical) sub-slices as parts of the same horizontal (vertical) slice, respectively.

#### 2. Scanning for critical points (Aggregation):

- Any sub-slice (3.2), or the vertical counterpart, is now treated as an individual time series to which we apply a single change-point estimate (2.3) with any  $\gamma \in [0, 1/2)$  as described in Section 2. (Also we tacitly assume the necessary modifications for vertical slices). Since we have  $n - N$  sub-slices for each  $i$ , we aggregate  $n - N$  estimated change point locations

$$\{\hat{u}(i, r) \in \mathbb{N}, \quad r = 1, \dots, n - N + 1\},$$

again, for any  $i$ . These locations will be called *critical points* in our spatial context.

- The locations  $\hat{u}(i, r)$  are integer-valued numbers since they are computed w.r.t. the  $(i, r)$ -th sub-slices. Hence, we need to map them back on our grid domain  $D$  via

$$\hat{U}(i, r) := (i, \hat{u}(i, r) + r - 1) \in D$$

for  $r = 1, \dots, n - N + 1, i = 1, \dots, m$ . For theoretical reasons, we will restrict the admissible change sets  $S$  by requiring  $(i, j) \notin S$  if  $j > n - N + 1$  or if  $i > m - N + 1$ . Also, for technical reasons, we have to set  $\hat{U}(i, r) := (i, 0)$  for  $r = n - N + 2, \dots, n$  and  $i = 1, \dots, m$ .

### 3. Selection of relevant critical points:

In this step, our aim is to identify the boundary  $B$  of the change set  $S$ . It will induce an estimate  $\hat{S}$  in a straightforward manner. Observe that only those critical points  $\hat{U}(i, r)$  which are adjacent to the change set  $S$  (or lie in  $B$ ) may help us to identify this boundary and therefore the set  $S$ . Asymptotically, i.e. as  $d \rightarrow \infty$  and with probability tending to 1, the points  $v \in B$  will correspond to those  $\hat{u}(i, r)$  and  $\hat{U}(i, r)$  that are based on correct estimation (2.6) and not on the spurious ones as in (2.7). Hence, we have to filter out the latter by selecting a set  $G$  of *relevant* critical points, based on  $\hat{U}$ , that is expected to be informative, based on suitable decision rules.

We present the *overlapping*  $(N, Q)$  rules, in form of a pseudocode. Let  $H(i)$  denote the set of relevant points w.r.t. the  $i$ -th horizontal slice and recall that we assume  $N \geq 4$  to be even. The overlapping  $(N, Q)$  rule is:

```

1: Choose some integer  $1 \leq Q \leq N - 2$ 
2: for  $i = 1$  to  $m$  do
3:    $H(i) \leftarrow \emptyset$ 
4:   for  $r = 1$  to  $n - N + 1$  do
5:     if  $\hat{U}(i, r) = \hat{U}(i, r + 1) = \dots = \hat{U}(i, r + Q)$  then
6:        $H(i) \leftarrow H(i) \cup \{\hat{U}(i, r)\}$ 
7:     end if
8:   end for
9: end for

```

The idea behind this algorithm is according to (2.6) and (2.7) that, assuming that the noise to change ratio lies below the threshold  $\mathcal{R}(\gamma, N)$ , only the following cases may occur for the estimate  $\hat{u}$  applied to (3.2):

- (a) There is a single change at  $1 \leq u < N - 1$ . In that case we know that asymptotically, as  $d \rightarrow \infty$ , this point is estimated correctly as a critical point with probability tending to 1.
- (b) There is more than one change point. This case will be excluded from our consideration (cf. conditions on the change sets (3.4) in Theorem 3.1 below).
- (c) There is no change in this sub-slice. Hence, asymptotically as  $d \rightarrow \infty$ , we estimate  $\hat{u} = \lfloor N/2 \rfloor$  spuriously with probability tending to 1.

For simplicity assume that  $Q = 1$  (the case  $Q > 1$  works in the same way). If some consecutive sub-slices, e.g. the  $(i, r)$ -th and  $(i, r + 1)$ -th, intersect the change set region  $S$ , such that both have a single change point, i.e. at  $u(i, r)$  and at  $u(i, r + 1)$ , then we are in case a) for both sub-slices and therefore

$P(\hat{U}(i, r+1) = \hat{U}(i, r)) \rightarrow 1$ , as  $d \rightarrow \infty$  which means that the condition of the 5th line, in the above algorithm, is fulfilled for  $Q = 1$ . On the other hand, if there is no change in at least one of the two subslices, we have  $P(\hat{U}(i, r+1) = \hat{U}(i, r)) \rightarrow 0$ , as  $d \rightarrow \infty$ , and the 5th line is always violated for any  $Q$ . Hence, the sets  $H(i)$ ,  $V(j)$  will asymptotically contain only points that correspond to change-points in the sub-slices.

The parameters  $N$  and  $Q$ , allow us to control the sensitivity. In particular, a smaller  $Q$  is less restrictive and therefore more sensitive, but less reliable. The overlapping rule is sketched in Figures 2 and 3 below. The illustration is based on a fragment of rectangular spatial observations of size  $m = 14$ ,  $n = 20$ . Each field corresponds to a point  $(i, j)$  in a straightforward manner. The sensitivity w.r.t. the parameters  $(N, Q)$  and  $\gamma$  is demonstrated in Figures 6-8, below.

Subsequently, we write  $V(j)$  for the set of relevant change points w.r.t. the  $j$ -th vertical slice. In case that we perform horizontal scanning only, we set formally  $V(j) = \emptyset$  for  $j = 1, \dots, n$  and analogously we set  $H(i) = \emptyset$  for  $i = 1, \dots, m$  if we would perform vertical scanning only.

The pooled set of all relevant critical points will be denoted by

$$G = H(1) \cup \dots \cup H(m) \cup V(1) \cup \dots \cup V(n).$$

#### 4. Connecting relevant critical points:

The set  $G$  should asymptotically contain only nodes adjacent to the boundary  $B$  of  $S$ . Hence, based on  $G$ , or on  $H(i)$  and  $V(j)$ , we try to identify as many nodes of  $S$  as possible. Here, this is carried out for each slice separately. (We tacitly restrict the class of change-sets  $S$  according to the Theorem 3.1). Let  $\hat{S}$  denote the estimate for  $S$ . The horizontal procedure is:

```

1:  $\hat{S} \leftarrow \emptyset$ 
2: for  $i = 1$  to  $m$  do
3:   if  $H(i) = \{(i, x_1), (i, x_2), \dots, (i, x_p)\}$ ,  $x_1 < x_2 < \dots < x_p$  with  $p \geq 2$ 
   then
4:      $\hat{S} := \hat{S} \cup \{(i, x_1 + 1), \dots, (i, x_p)\}$ 
5:   end if
6: end for

```

and the analogous vertical procedure is:

```

1: for  $j = 1$  to  $n$  do
2:   if  $V(j) = \{(x_1, j), (x_2, j), \dots, (x_p, j)\}$ ,  $x_1 < x_2 < \dots < x_p$  with  $p \geq 2$ 
   then
3:      $\hat{S} := \hat{S} \cup \{(x_1 + 1, j), \dots, (x_p, j)\}$ 
4:   end if
5: end for

```

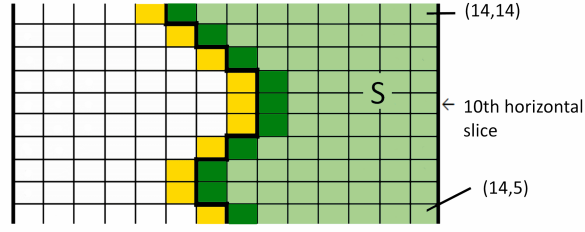


Figure 2: The common change set  $S$  is indicated by green color. The darker fields indicate the boundary of  $S$ . The yellow fields indicate relevant critical points that will be eventually selected, i.e. with probability tending to 1 as  $d \rightarrow \infty$ , applying the overlapping  $(4, 2)$  rule with any  $\gamma \in [0, 1/2)$ .

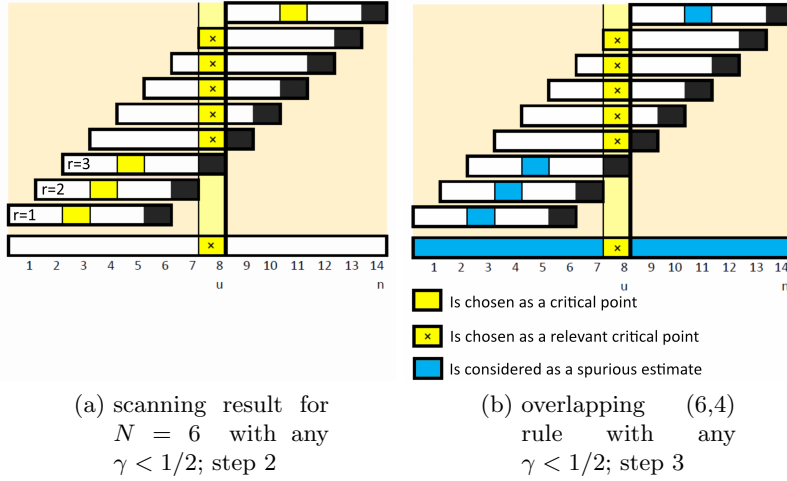


Figure 3: The figures illustrate the algorithm for the overlapping rule. All figures show the 10-th horizontal slice of Figure 2 and the corresponding  $n - N$  overlapping sub-slices. It is indicated which points are selected as relevant critical points asymptotically with probability tending to 1 as  $d \rightarrow \infty$

**Theorem 3.1.** Assume a rectangular domain  $D$  with  $\min\{m, n\} \geq \xi \geq 4$ , a boundary  $B$  and a common connected change set  $S$  with  $d(S, B) \geq \xi - 1$  for some  $\xi \in \mathbb{N}$ . Define vertical and horizontal intersections by

$$\begin{aligned} H_i &:= S \cap \{(i, l) \mid l = 1, \dots, n\}, \\ V_j &:= S \cap \{(l, j) \mid l = 1, \dots, m\}, \end{aligned} \quad (3.4)$$

for all  $j = 1, \dots, n, i = 1, \dots, m$ , respectively. Furthermore, Assume that all  $H_i$  and  $V_j$  are either empty or connected sets. We have to state three different assumptions:

1. Assume that we use the horizontal approach and that for all  $i = 1, \dots, m$  it holds that  $|H_i| \geq \xi$  if  $H_i \neq \emptyset$ .
2. Assume that we use the vertical approach and that for all  $j = 1, \dots, n$  it holds that  $|V_j| \geq \xi$  if  $V_j \neq \emptyset$ .
3. Assume that we combine the horizontal together with the vertical approach and that for all  $i = 1, \dots, m, j = 1, \dots, n$  it holds that  $\max\{|H_i|, |V_j|\} \geq \xi$  if  $(i, j) \in H_i \cap V_j$ .

Assume that we use the overlapping  $(N, Q)$  rule, with  $4 \leq N \leq \xi$  and  $1 \leq Q \leq N - 2$  where  $N$  is even. Furthermore, let  $\gamma \in [0, 1/2)$  and the ratio  $\rho$  be below  $\mathcal{R}(\gamma, \xi)$ . Under either of the above Assumptions 1-3, given that (2.5) holds true w.r.t. all sub-slices (2.1), it holds that, as  $d \rightarrow \infty$ ,

$$P(\hat{S} = S) \rightarrow 1. \quad (3.5)$$

*Proof.* Conditions on  $H_i$  and  $V_j$  ensure that only two cases may occur. Either a sub-slice contains a single change point or does not contain a change point at all. The number of sub-slices is fixed and finite. Hence, the overall consistency of  $\hat{S}$  follows from the consistency of all estimates  $\hat{u}(i, r)$  in case of a change and from the fact of spurious estimation when there is no change (cf. Torgovitski (2015, Theorems 2.6, 2.13 and Remark 2.17)).  $\square$

## 4 Simulations

We start this section by illustrating the estimation procedure and the corresponding Theorem 3.1 of the previous Section 3.

We begin with parameters that will be common in our simulations. For simplicity, we consider a domain  $D$  with  $m, n = 100$  and assume the noise to be i.i.d. normally distributed with  $\varepsilon_1(1, 1) = N(0, \sigma^2)$ . We consider a single common change set scenario (see Section 1 and 2) and define test change sets  $S$  with radius  $w$ , centered at a point  $v \in D$ , by

$$S_{w,v} := \{u \in D \mid \|u - v\|_p \leq w\}.$$

Here,  $\|u - v\|_p$  is the usual  $p$ -Norm for vectors in  $\mathbb{R}^2$  and  $p = \infty$  denotes the maximum norm. We call such sets *rectangular-shaped* for  $p = \infty$ , *round-shaped* for  $p = 2$  and *diamond-shaped* for  $p = 1$ . The reference mean level  $m_k(S^c)$  is set to  $m_k = k$  for  $k = 1, \dots, d$ .

*Remark 4.1.* Recall, that  $d_J$  denotes the Jaccard distance defined in (1.1). Clearly, relation (3.5) implies  $P(d_J(\hat{S}, S) = 0) \rightarrow 1$  as  $d \rightarrow \infty$  which in turn yields  $E(d_J(\hat{S}, S)) \rightarrow 0$  as  $d \rightarrow \infty$ . The latter follows e.g. due to uniform integrability of  $d_J(\hat{S}, S) \in [0, 1]$ . In our simulations we demonstrate the influence of various parameters on the expected Jaccard distance  $Ed_J = E(d_J(\hat{S}, S))$  which is approximated based on 100 repetitions.

Table 1 shows  $Ed_J$  for the overlapping  $(N, Q)$  rules w.r.t. different parameters  $d, N, Q$  and  $\gamma$ . Generally, it is not clear which combination of sensitivity parameters  $(N, Q)$  and  $\gamma$  is preferable. Hence, our advice is to plot different estimates and to rely on visual inspection (cf. Figures 6 - 8 below). Nevertheless, we see two tendencies where either the expected distance  $Ed_J$  improves for larger  $d$ , e.g. for  $N = 4, Q = 1$  and  $\gamma = 0.3$ , or worsens, e.g. for  $N = 4, Q = 1$  and  $\gamma = 0$ . In accordance with the theory, the former happens if the ratio  $\rho$  is below the threshold  $\mathcal{R}$  and the latter when  $\rho$  is above. Notice, that the precision does not monotonously increase in  $\gamma$ .

The Figures 6 - 8 are based on a spatio-temporal sequence which is illustrated in Figure 5. Comparing the Figures 6 and 7 for  $\gamma = 1/4$  we see that a larger  $d$  improves the estimation. Clearly, a smaller  $Q$  yields more sensitive estimates but on the other hand larger parameters  $Q$  may isolate the change sets better. Table 1 and Figures 6-8 show that the usage of the horizontal procedure together with the vertical procedure might be better or worse than the plain horizontal approach. However, for some change sets, e.g. diamond-shaped, it is necessary to use both directions in order to obtain a consistent estimate (cf. Figure 4).

	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
<b>overlapping (4,1)</b>					
$d = 100$	0.46 (0.51)	0.47 (0.54)	0.49 (0.54)	0.50 (0.55)	0.51 (0.55)
$d = 200$	0.68 (0.52)	0.44 (0.46)	0.45 (0.53)	0.49 (0.55)	0.51 (0.55)
$d = 300$	0.89 (0.79)	0.52 (0.38)	0.41 (0.49)	0.48 (0.54)	0.51 (0.55)
$d = 500$	0.97 (0.94)	0.72 (0.52)	0.30 (0.28)	0.43 (0.53)	0.50 (0.55)
$d = 1000$	0.99 (0.99)	0.82 (0.66)	0.24 (0.06)	0.23 (0.35)	0.49 (0.54)
<b>overlapping (4,2)</b>					
$d = 100$	0.92 (0.85)	0.74 (0.59)	0.53 (0.47)	0.44 (0.50)	0.45 (0.53)
$d = 200$	0.99 (0.98)	0.92 (0.85)	0.66 (0.48)	0.42 (0.43)	0.42 (0.52)
$d = 300$	1.00 (0.99)	0.95 (0.91)	0.73 (0.54)	0.38 (0.33)	0.41 (0.50)
$d = 500$	1.00 (1.00)	0.97 (0.94)	0.75 (0.56)	0.28 (0.16)	0.36 (0.47)
$d = 1000$	1.00 (1.00)	0.98 (0.97)	0.67 (0.45)	0.12 (0.02)	0.23 (0.35)
<b>overlapping (6,2)</b>					
$d = 100$	0.42 (0.48)	0.42 (0.51)	0.43 (0.52)	0.45 (0.53)	0.46 (0.54)
$d = 200$	0.33 (0.36)	0.36 (0.45)	0.40 (0.50)	0.43 (0.53)	0.46 (0.54)
$d = 300$	0.24 (0.20)	0.27 (0.36)	0.36 (0.48)	0.41 (0.52)	0.45 (0.54)
$d = 500$	0.10 (0.04)	0.11 (0.16)	0.26 (0.38)	0.38 (0.50)	0.44 (0.53)
$d = 1000$	0.01 (0.00)	0.01 (0.01)	0.07 (0.12)	0.28 (0.41)	0.42 (0.52)
<b>overlapping (6,4)</b>					
$d = 100$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.95 (0.89)	0.65 (0.46)
$d = 200$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.95 (0.90)	0.53 (0.30)
$d = 300$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.94 (0.89)	0.40 (0.18)
$d = 500$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.93 (0.86)	0.23 (0.06)
$d = 1000$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.90 (0.82)	0.05 (0.00)

Table 1: The expected Jaccard distance  $Ed_J$  for the overlapping algorithm for the horizontal (horizontal + vertical) approach. For simplicity we set  $m_k(S) = k + (-1)^k$ ,  $k = 1, \dots, d$ , i.e.  $m_k(S) = m_k(S^c) + (-1)^k$ , and  $\sigma^2 = 2$ . The change set  $S_{w,v}$  is rectangular-shaped centered at  $v = (50, 50)$  with  $w = 100/3$

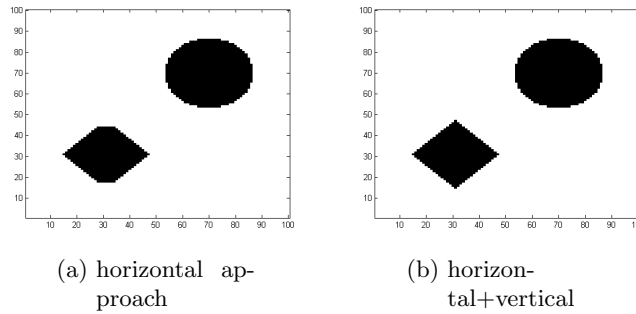


Figure 4: The figure shows the overlapping (16,8) estimate with  $\gamma = 0$  for multiple diamond- and round-shaped change sets  $D = S_D + S_R + S^c$  with  $w = 100/6$ . The horizontal procedure is consistent for the round-shaped set but only the horizontal + vertical approach yields a consistent estimate for the diamond-shaped change set. It holds  $m_k(S_D) = m_k(S^c) + (-1)^k$ ,  $m_k(S_R) = m_k(S^c)$ ,  $k = 1, \dots, d$  and  $d = 1000$  with  $\sigma^2 = 1$

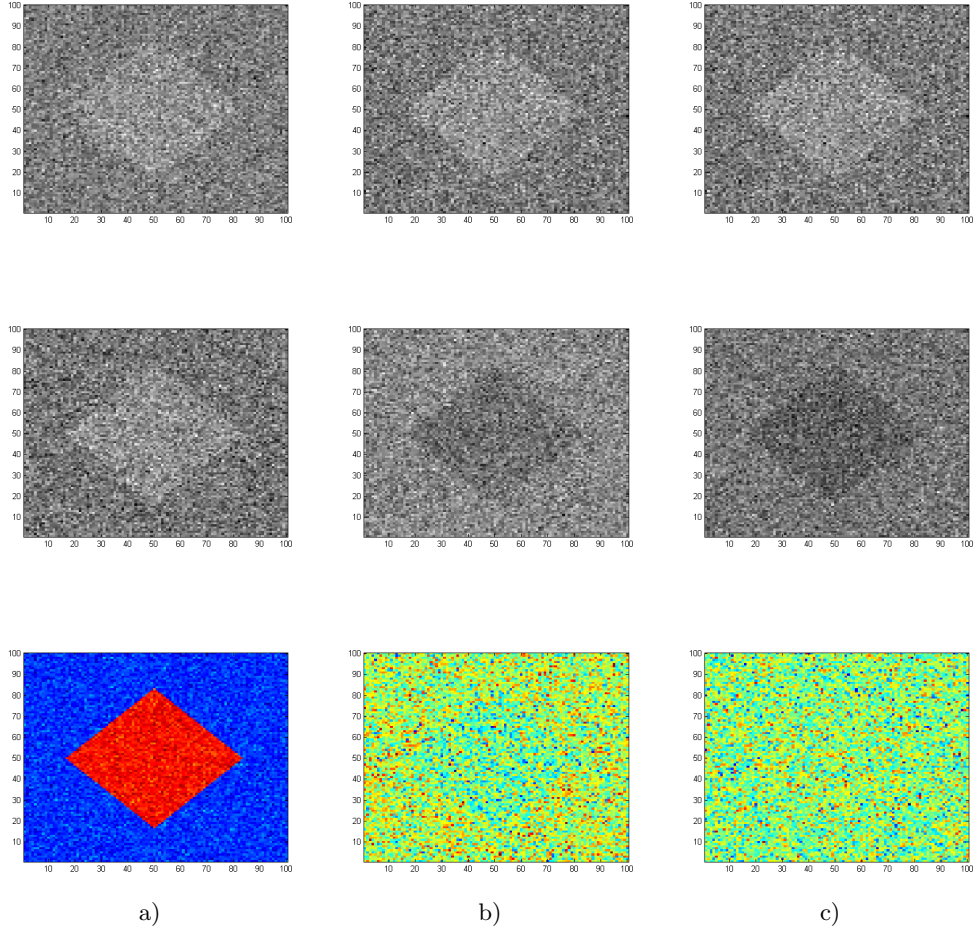


Figure 5: The first two rows show the observations  $X_1, \dots, X_6$  with  $\sigma^2 = 2$  for  $m_k(S) = k + (-1)^k$  and  $m_k(S^c) = k$ . The third row shows the averages  $\bar{X}_{500}$ . In a) for the simple case of  $m_k(S) = 0$  and  $m_k(S^c) = 1$ . In b) for the case of  $m_k(S) = 0$  and  $m_k(S^c) = (-1)^k$  and finally in c) for  $m_k(S) = k$  and  $m_k(S^c) = k + (-1)^k$ . Notice that by averaging we loose (or at least do not gain) information in the last two settings

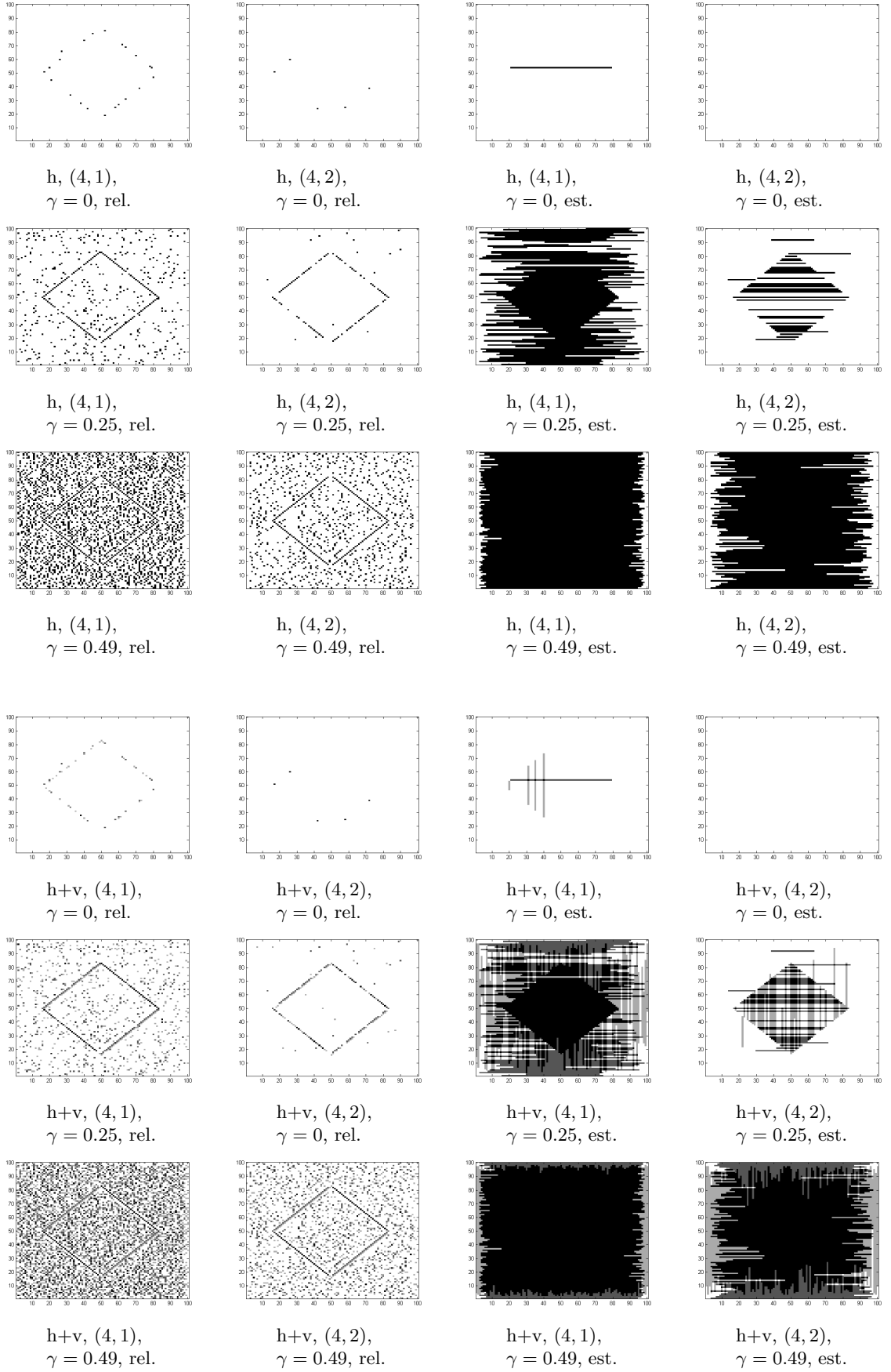


Figure 6: The figures show the relevant critical points (rel.) and the corresponding estimates (est.) selected by the horizontal (h) or the horizontal+vertical (h+v) approach using the overlapping  $(N, Q)$  rule and based on a diamond-shaped change set. It is  $m_k(S) = k + (-1)^k$ ,  $k = 1, \dots, d$  and  $d = 500$  with  $\sigma^2 = 2$

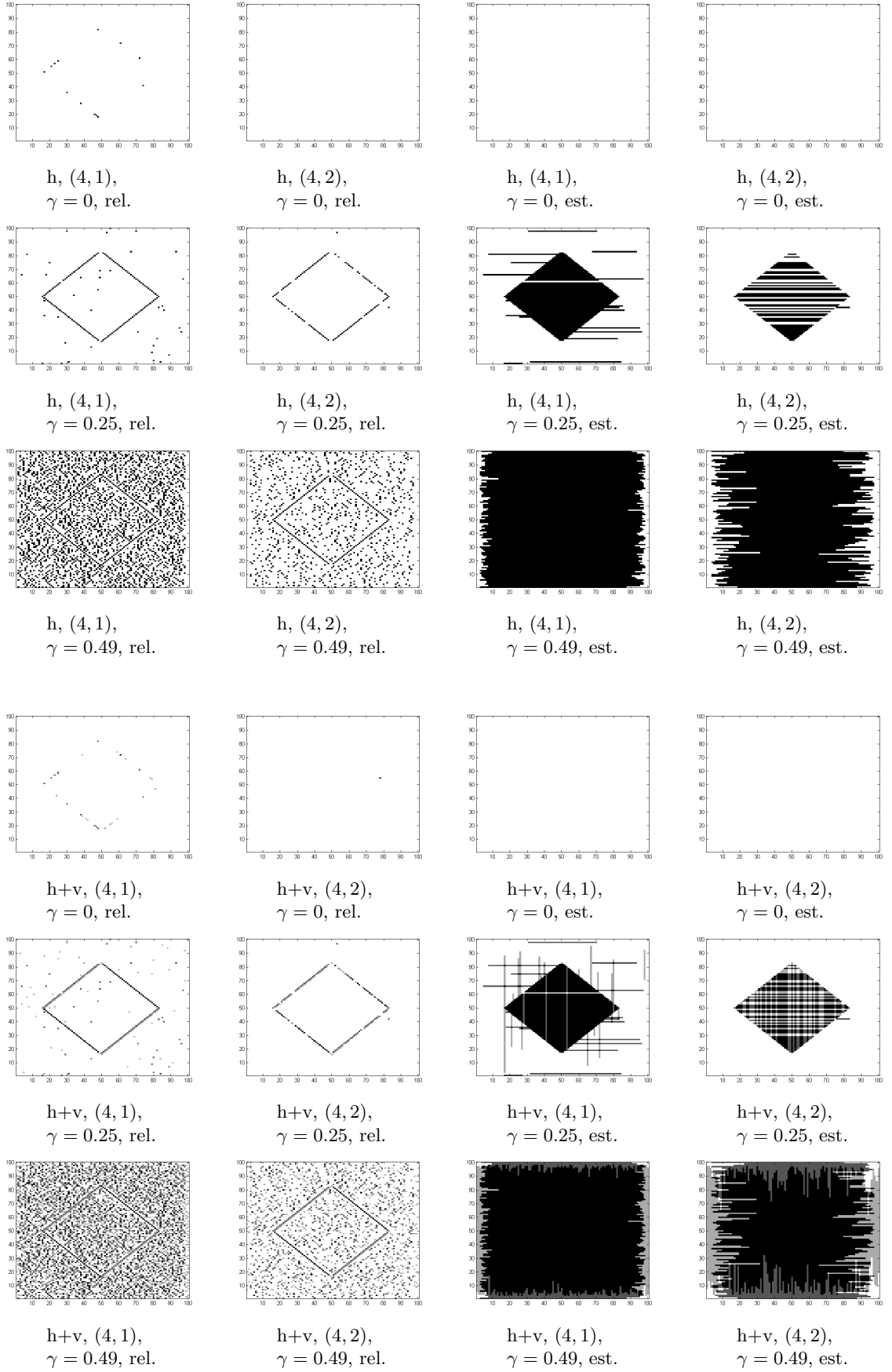


Figure 7: The figures show the relevant critical points (rel.) and the corresponding estimates (est.) selected by the horizontal (h) or the horizontal+vertical (h+v) approach using the overlapping  $(N, Q)$  rule and based on a diamond-shaped change set. It is  $m_k(S) = k + (-1)^k, k = 1, \dots, d$  and  $d = 1000$  with  $\sigma^2 = 2$

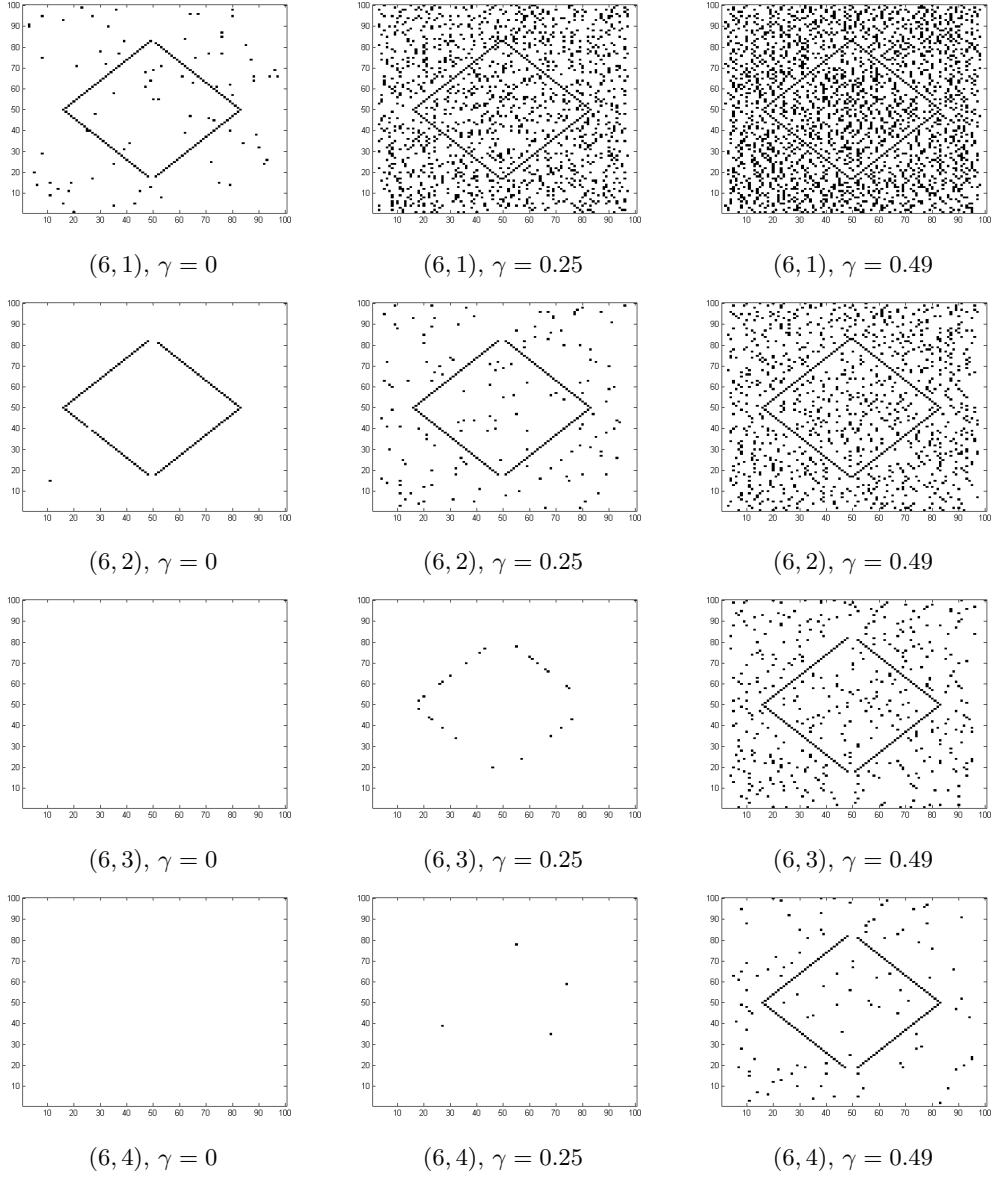


Figure 8: The figures show the relevant critical points selected by the horizontal approach using the overlapping  $(N, Q)$  rule and based on a diamond-shaped change set. It is  $m_k(S) = k + (-1)^k$ ,  $k = 1, \dots, d$  and  $d = 1000$  with  $\sigma^2 = 2$

## References

- Arnold M, Wied D (2012) Testing for structural change in spatial regions at unknown positions. Discussion Paper 19/2012, SFB 823.
- Arnold M, Raabe N, Wied D (2014) Identifying different areas of inhomogeneous mineral subsoil: spatial fluctuation approaches. Communications in Statistics - Simulation and Computation. doi:10.1080/03610918.2013.861487
- Bai J (2010) Common breaks in means and variances for panel data. Journal of Econometrics 157:78–92
- Bleakley K, Vert J P (2010) Fast detection of multiple change-points shared by many signals using group LARS. Neural Inform Process Syst 22:2343–2352
- Bleakley K, Vert J P (2011a) The group fused LASSO for multiple change-point detection. arXiv:1106.4199v1
- Bleakley K, Vert J P (2011b) The group fused LASSO for multiple change-point detection. Technical report HAL-00602121
- Hadri K, Larsson R, Rao Y (2012) Testing for stationarity with break in panels where the time dimension is finite. Bull Econ Res, Issue Supplement s1. 64:s123–s148
- Kim D (2014) Common breaks in time trends for large panel data with a factor structure. Econom J 17:301–337
- Polzehl J, Spokoiny V (2000) Adaptive weights smoothing with applications to image restoration. J R Statist Soc B 62:335–354
- Qiu P (2007) Jump surface estimation, edge detection, and image restoration. Journal of the American Statistical Association 102(478):745–756
- Torgovitski L (2015) Panel data segmentation under finite time horizon. arXiv:1501.00177v2